AMENDMENTS

In the Specification:

Please replace the Field of the Invention section with the following replacement section in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

Field of the Invention

The present invention relates to a synthetic gene coding for human granulocyte colony stimulating factor (hG-CSF) which enables expression in *E. coli* with an improved expression level, enabling an expression level ~~being~~ equal to or higher than 52% of the recombinant hG-CSF to the total proteins after expression.

hG-CSF belongs to a family of stimulating factors which regulate the differentiation and proliferation of hematopoetic mammalian cells. They have a major role in ~~the~~ neutrophil formation and are therefore suitable for use in medicine in the field of hematology and oncology.

Two forms of hG-CSF are currently available for clinical use on the market: lenograstim, which is glycosylated and is obtained by the expression in mammalian cell line; and filgrastim, which is non-glycosylated and is obtained by the expression in the bacterium *Escherichia coli* (*E. coli*).

Please replace the last full paragraph on page 1 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The GC rich regions also have an impact on the translational efficiency in *E. coli* if a stable double stranded RNA is formed in the mRNA secondary structure. This impact is the highest when the GC rich regions of mRNA are found either in the RBS, or in the direct proximity of the RBS or also in the direct proximity of the start codon (Makrides SC, Microbiological Reviews, 60:512-538 (1996); Baneyx F, Current Opinion in Biotechnology, 10:411-421 (1999)).

Please replace the first full paragraph on page 2 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

There are known several methods for the prediction of the secondary structure and calculating minimal free energy of individual RNA molecule which is supposed to be the basic rule for the most stable/most probable structure (SantaLucia J Jr and Turner DH, Biopolymers, 44:309-319 (1997)). The reliable algorithms for the prediction of the correct secondary structure are not known, with the exception of some cases. There has been no evidence for the quantitative correlation with the expression level (Smit MH and van Duin J., J. Mol. Biol., 244, 144-150 (1994)). It is still impossible to predict the tertiary structures of RNA (Tinoco, I. and Bustamante C., J. Mol. Biol, 293:271-281 (1 999)).

Please replace the paragraph bridging pages 2 and 3 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The expression of hG-CSF in *E. coli* with the yield up to 17% of hG-CSF to total cellular bacterial proteins is described in Devlin PE et al, Gene 65:13-22 (1988). Such yield was reached with partial optimization of DNA sequence in the 5' end of the G-CSF coding region (codons coding for the first four amino acids) whereby the GC region was replaced with an AT region and a relatively strong lambda phage promoter was used. This expression level is not very high ~~what~~ which leads to lower production yields and is less economical in the large-scale production.

Please replace the first full paragraph on page 3 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The use of a synthetic gene and the expression level of about 30% are described in Kang SH et al, Biotechnology letters, 17(7):687-692 (1995). This level was attained by the introduction of *E. coli* preference codons, by the modifications in the TIR region, and with the additional modifications of codon sets whereby the 3' end of the gene was not essentially

changed. Thus, for attaining the stated expression level the changes of the gene in the TIR region were needed and the expression level did not exceed 30%.

Please replace the second full paragraph on page 3 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The patent US5840543 describes the synthetic gene coding for hG-CSF which was constructed by the introduction of AT rich regions at the 5' end of the gene and with the replacement of *E. coli* rare codons with *E. coli* preference codons. Under the control of the Trp promoter, ~~the~~ expression with the yield of 11% hG-CSF to total cellular proteins was reached. On the other hand, the addition of leucine and threonine or their combination into the fermentation medium (where the bacteria were cultivated) led to the accumulation of up to 35% of hG-CSF regarding total cellular proteins. Such expression level was therefore reached by the addition of amino acids into the fermentation medium ~~what~~ which is an additional cost in the process for production of hG-CSF and is not economical for the industrial production. Only optimization of the gene coding for hG-CSF did not enable a higher expression level of hG-CSF.

Please replace the first full paragraph on page 4 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

Until now, there has been no entirely developed rule known on how to combine ~~the~~ codons in order to obtain ~~the~~ secondary or tertiary mRNA structures which are optimal for expression. Although there exist some mathematical and structural models for predicting and thermodynamic~~al~~ stability of secondary structures, ~~but~~ they are too unreliable to predict the secondary structures. On the other hand, there are no such models for predicting the tertiary structures. These currently accessible models therefore do not enable the prediction of the impact of the codons on the expression level.

Please replace the second full paragraph on page 4 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

There are no reports in either the patent or the scientific literature on the more efficient way for solving the problem of low expression level of the ~~native~~ gene coding for hG-CSF in *E. coli*.

Please replace the fourth full paragraph on page 4 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The above and other objects ~~is solved~~ are addressed by a DNA sequence according SEQ ID NO:1~~to claim 1~~, and by a process for the construction of such a DNA sequence as disclosed herein ~~according to claim 15~~. The present invention also provides an expression plasmid ~~according to claim 6 or 7~~, an expression system ~~according to claim 11 or 12~~, a process for the expression of hG-CSF ~~according to claim 20~~, and a process for the manufacture of a pharmaceutical composition ~~according to claim 24~~. Preferred embodiments of these and other aspects of the invention are also disclosed ~~defined in sub-claims~~ herein.

Please replace the paragraph bridging pages 4 and 5 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

A ~~The~~ significant feature of the present invention is that the use of a synthetic gene coding for hG-CSF enables one to attain an expression level (accumulation) in *E. coli* being equal to or higher than 52% of recombinant hG-CSF regarding the total proteins in *E. coli*- Preferably, an expression plasmid containing a strong T7 promoter is used for the expression. The synthetic gene coding for hG-CSF is constructed by using a complex combination of two methods which enable the construction of an optimized synthetic gene (coding for hG-CSF) for its expression in *E. coli*. The first method includes the replacement of some rare *E. coli* codons which are unfavorable for expression in *E. coli* by *E. coli* preference codons for which are more favorable for ~~the~~ expression in *E. coli*. The second method includes the replacement of some GC rich regions by AT rich regions. Some parts of the synthetic gene of the present

invention are constructed by using one of the two methods, for some parts the combination of the two methods is used, whereas some parts of the gene are not changed. In the construction procedure of the synthetic gene coding for hG-CSF, which is also the subject of the present invention, the non coding (5'-untranslated) regions are preferably not changed. Advantageously, this means that there are no modifications in either the translation initiation region (TIR) or in the ribosome binding site (RBS), or in the region between the start codon and RBS.

Please replace the third full paragraph on page 6 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The term 'hG-CSF', as used herein, refers to human granulocyte-colony stimulating factor, comprising the recombinant hG-CSF obtained by ~~the~~ expression in *E. coli*.

Please replace the fourth full paragraph on page 7 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The term 'synthetic gene', as used herein, refers to the gene prepared from short double stranded DNA fragments which are composed of synthetic complementary oligonucleotides. This synthetic gene differs from the native gene (e.g., cDNA) only in the nucleotide sequence, while ~~whereby~~ the amino acid sequence remains unchanged. The synthetic gene is obtained by the techniques of ~~the~~ recombinant DNA technology.

Please replace the first full paragraph on page 9 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

After the initial preparation of the native gene coding for hG-CSF and of plasmids, the optimization of the native gene coding for hG-CSF is performed. This means that the synthetic gene coding for hG-CSF is constructed. The optimization begins with the division of the native gene coding for hG-CSF into four (I, II, III, and IV) segments, which are or will be separated with single restriction sites after the oligonucleotide mutagenesis.~~ and in the individual segments the changes are introduced.~~ Changes are introduced in individual segments. In some individual segments, ~~the~~ changes in the gene sequence are introduced, whereas in certain segments the gene is not changed (Figure 1). The ~~obtained~~ optimized synthetic gene coding for hG-CSF therefore consists of a partially preserved native sequence (segment 111) and of 5' and 3' coding regions which are synthesized de novo (segments I, II, and IV).

Please replace the second full paragraph on page 11 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

The fermentation process may be performed in the presence or in the absence of the antibiotic that corresponds to <u>the</u> resistance gene ~~which is~~ inserted into the plasmid vector, e.g. with ampicilline or kanamycin at an appropriate concentration or in the absence thereof. It has <u>also</u> been found that the fermentation and thus the accumulation of hG-CSF was highly effective ~~also~~ without a selection pressure.

Please replace the paragraph bridging pages 13 and 14 with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through.

<u>Example 1b: Codon optimization (Figure 1)</u>

In <u>a</u> ~~the~~ first optimization step<u>,</u> the synthetic gene between the restriction sites NdeI and SacI was constructed by ligation of five cassettes (A, B, C, D, E) which were composed of complementary oligonucleotides. This synthetic part of the gene represents the segment I. With the segment I<u>,</u> the part of the native gene for hG-CSF between the restriction sites NdeI and SacI was replaced. This was performed by the excision of the first part of the gene between the restriction sites NdeI and SacI<u>,</u> and ~~its~~ replacement with the synthetically prepared cassette. The process was performed in two steps. In the first step, the cassette A was ligated to the NdeI site and the cassette E was ligated to the SacI site. After 16 hours at 16°C the ligation mixture was precipitated with ethanol to remove the excess of ~~(not bound)~~ <u>unbound</u> oligonucleotides. In <u>a</u> ~~the~~ second ~~steps~~ <u>step</u> the central part of the whole cassette (cassettes B, C and D) from the three previously ligated complementary oligonucleotides was added and the ligation was performed for 16 hours at 16°C.

Please replace the first full paragraph on page 14 of the specification with the following replacement paragraph in which subject matter added is indicated by underlining and subject matter deleted is indicated by strike-through:

In <u>a</u> ~~the~~ second optimization step<u>,</u> ~~the~~ two ~~for E. coli most~~ critical codons<u> for *E. coli*,</u> located in ~~the~~ segment III, namely, CGG→CGT (Arg148) and GGA→GGT (Gly150), were replaced by

using ~~the~~ oligonucleotide-directed mutagenesis (TransformerTM Site-Directed Mutagenesis Kit (Clontech)).